METHODOLOGICAL INNOVATIONS Online

Methodological Innovations Online (2011) 6(2) 52-70

Selecting cases for in-depth study from a survey dataset: an application of Ragin's configurational methods[1]

Judith Glaesser[a] & Barry Cooper[a]

[a]School of Education, Durham University

Abstract

While 'establishing the phenomena', to use Merton's phrase, is an important part of the sociological enterprise, in then accounting for such empirical regularities, theoretical models are required to understand causal processes. Both regression analysis and configurational methods applied to large datasets can establish patterns of relationships. Following a realist view, we assume that causal mechanisms have generated such patterns, and sound theoretical models are required to understand them. In-depth case studies can contribute to advancing such causal knowledge. We describe how, in the particular context of the configurational mode of analysis that characterises Ragin's Qualitative Comparative Analysis (QCA), we have selected individuals for in-depth study with the eventual purpose of advancing causal or explanatory understanding of conjunctural empirical regularities concerning educational careers. While forms of regression analysis seek to establish the net effects of 'independent' variables, QCA, employing Boolean algebra, analyses the conjunctions of conditions sufficient and/or necessary for an outcome to occur. QCA aims to preserve, holistically, the features of cases and is therefore well-suited to case selection. We use QCA both to undertake an initial large scale cross-case analysis and to subsequently select cases to develop theoretical understanding via within-case analysis. Using QCA's measures of consistency with relations of sufficiency and necessity, we can classify cases as typical and deviant, with these two types of cases playing different roles in testing and developing theory. Drawing on analyses of the German SOEP dataset undertaken as part of a larger study which is applying case-based configurational methods to English and German survey datasets while undertaking subsequent in-depth interviews with selected cases, we demonstrate how QCA can be used to select cases for interview in a systematic and theoretically informed manner.

Keywords: Case selection, case study, causal mechanisms, Configurational Analysis, Qualitative Comparative Analysis (QCA)

Introduction

Sociological work takes many forms. Considering just the social survey, a researcher may be aiming simply to describe some set of affairs (such as the distribution of educational achievement over categories of social class origin at various points of time), or be aiming to use his or her analysis to make predictions about cases outside the analysed dataset or, counterfactually, about what would occur were some characteristic of a studied case to be changed via social policy[2], or be aiming to generate theory inductively[3], or be aiming to develop and/or test existing theory of various types, or be aiming to develop causal knowledge of some phenomenon. We have no doubt that the difficulty of producing valid work rises as the researcher moves from

---

Correspondence: Dr Judith Glaesser, Prof Barry Cooper, School of Education, Durham University, Leazes Road, Durham, DH1 1TA. Tel.: +44(0)191 3348308. E-mail: Judith.Glaesser@durham.ac.uk Barry.Cooper@durham.ac.uk

ISSN: 1748-0612online
DOI: 10.4256/mio.2010.0035

a descriptive to a causal focus, though we would not wish to underestimate the difficulty of producing valid description per se. It is, however, the selection of case studies specifically to support the move from a descriptive to a causal focus that we wish to address here. While 'establishing the phenomenon' (Merton, 1987) is an important part of the sociological enterprise, in then accounting for such empirical regularities, theoretical understanding of some kind is demanded. As Goldthorpe (2007a, especially chapters 6 and 9) has cogently argued, what is required to understand the empirical regularities produced via quantitative data analysis are substantive theoretical models of the processes that have generated them. The same argument has been made under the banner of realism (e.g. Pawson, 1989). We find these arguments compelling and wish to describe how, in the particular context of the configurational mode of analysis that characterises Charles Ragin's Qualitative Comparative Analysis (QCA), we have selected individuals for in-depth study with the eventual purpose of advancing causal or explanatory understanding of conjunctural empirical regularities concerning educational careers. To do so, we first discuss the relationship of large n analysis and case studies, followed by a brief introduction to QCA. Referring briefly to previous approaches to case selection, we then describe and discuss our own approach.

We have been using QCA for some time to produce configurational analyses of the relationships between various conditions and outcomes in large survey datasets (Glaesser, 2008, Cooper & Glaesser, 2008, Cooper, 2005[4]). While it was originally developed for use with small to medium-sized datasets, Ragin himself has applied QCA to a large dataset, the so-called bell-curve data, comparing the results with those of a logistic regression analysis (Ragin 2006).

The details of QCA, which favours configurational accounts over the 'net effects' models produced by regression analyses, will be addressed later. The key point is that the results of such configurational analyses are expressed as summaries of the various pathways to some outcome, showing, in set theoretic notation, the different combinations of some conditions that are sufficient, or quasi-sufficient, for the outcome to occur. However, before discussing the differences between QCA-based and regression-based analyses of large datasets, we must acknowledge a feature of both approaches noted previously by Pawson (2008) since this has motivated our decision to employ in-depth case studies alongside QCA in our most recent work. The problem is the well-known one concerning, in the variable analytic tradition, the relation between correlation and causation. A regression analysis of educational achievement, for example, might produce a summary equation describing a strong relationship between the independent variables capturing familial socio-economic status, general cognitive ability, sex and the dependent variable, measured perhaps by number of passes achieved in an examination taken in late adolescence. The problem remains, however, of whether these independent variables are the real causes of the outcome and, if they are, how and why they are. The same problem arises with configurational analyses. There are several well-known aspects to this problem (Morgan and Winship, 2007). We will mention here just two that have motivated our decision to employ case studies alongside QCA analyses of large scale survey data.

1. Variables like SES, general cognitive ability and sex summarise in one measure complexes of features of individuals, their contexts and relationships. For this reason different values of these variables cannot easily be seen as simple sociological equivalents of, for example, the masses a physicist might hang from a spring in an experiment designed to model the relation between varying weights and the consequential varying extensions of the spring. Even if we were to accept, therefore, as a *plausible* causal conclusion based on quantitative analysis, that 'sex' makes a difference to achievement, we would often want to know, in more detail, what it was about 'sex', in the form of gender relationships, decision-making, etc., that actually caused the differences captured in our equations. The same point can be made about these sorts of factors when they appear in conjunctural analyses.

2. Even in cases where the independent variables are simpler in form, the usual problems of spuriousness have to be addressed. In a regression analysis, variable A may be found to strongly predict variable C, but the real cause of C might be the unmeasured variable B, that precedes A and happens to be strongly correlated with A. Or both A and C may be causally dependent on D, with this being the real explanation of the association between A and C. More height (a simple variable), for example, might be found to predict more educational achievement, but it would be likely that this relationship was not

> causal but spurious. Clearly, within a conjunctural approach, similar problems can arise when a conjunction of two factors is found to predict some outcome.

Taking regression analysis as an example, in both 1 and 2 above some researchers might be willing to stop, either (in the case of 1) having more fully specified the component independent variables comprising something like SES or (in the case of 2) having shown that the original relationship was spurious and identified the omitted independent variable. Such researchers are likely to share the view of King et al. (1994, p. 86) that talk of mechanisms and processes can be reduced (indeed, for these authors, must be reduced) to talk of yet more intervening variables and the establishment of directed relationships between them[5]. A researcher undertaking conjunctural analyses could take a parallel position.

However, Bhaskar's work (e.g. 1978), alongside that of many others, has made it quite clear that the existence of a simple regularity is neither sufficient nor necessary evidence for the existence of some causal mechanism, where the latter is understood in terms of, for example, structural relations between agents with varying properties. Given these considerations, social scientists often, implicitly or explicitly, work towards specifying types, or kinds, of individuals, institutions or societies, seeing these as entities with specific dispositions, preferences and causal powers. Recent writing on typological theorising and process-tracing (George & Bennett, 2005) and explanatory typologies (Elman, 2005) has much good advice to offer in this regard.

We would not want, as some might, to argue that a case-based perspective rules out a concern with documenting empirical regularities (though, as we noted above, as in the example of regression, these cannot be assumed to demonstrate causality directly). However, since various causal properties and tendencies can be enabled and/or blocked by other causal powers operating in any social context, we can, on this view, at best expect to find complex and/or less than perfect regularities when analysing large scale datasets, even before the introduction of any additional role for chance or contingency. This view of causation is summarised well by Little (1997) in his post-positivist approach to social science:

> [This approach] recognizes that there is a degree of pattern in social life—but emphasizes that these patterns fall far short of the regularities associated with laws of nature. It emphasizes contingency of social processes and outcomes. It insists upon the importance and legitimacy of eclectic use of social theory: the processes are heterogeneous, and therefore it is appropriate to appeal to different types of social theories as we explain social processes. It emphasizes the importance of path-dependence in social outcomes. It suggests that the most valid scientific statements in the social sciences have to do with the discovery of concrete social-causal mechanisms, through which some types of social outcomes come about. And finally, it highlights what I call 'methodological localism': the insight that the foundation of social action and outcome is the local, socially-located and socially constructed individual person. The individual is *socially constructed*, in that her modes of behavior, thought, and reasoning are created through a specific set of prior social interactions. And her actions are *socially situated*, in the sense that they are responsive to the institutional setting in which she chooses to act. Purposive individuals, embodied with powers and constraints, pursue their goals in specific institutional settings; and regularities of social outcome often result. (Little, 1997, emphasis in original)

Considering the sort of empirical regularities that are typical of work on social class and education, such as the correlation between social class of origin and eventual academic achievement, sociologists – and, for that matter, laypeople – do not find it hard to create plausible actor-centred causal narratives that might account for the observed (if imperfect) regularities. The two main sociological competitors are, of course, those due to Bourdieu and Boudon. While the latter's rational action account, especially in the hands of Goldthorpe (2007b), tends to play down differences between actors, focusing instead on the differing contexts of costs and benefits they act within to explain class differences in outcome, Bourdieu's use of *habitus* explicitly focuses on the differences in actors' dispositions. Given their mutual focus on action, we would also note that both Boudon's and Bourdieu's key theoretical ideas, of primary and secondary effects and of capitals / habitus respectively, have an affinity with case-based approaches. They also have an affinity with theoretical statements employing the concepts of sufficient and/or necessary conditions, i.e. the building blocks of the set theoretic configurational approach developed by Ragin in QCA:

- Boudon (1974a) accounts for the social distribution of educational achievement in terms of primary and secondary effects. The primary effects of social class create some part of the differences in measured ability/achievement early in a child's career while secondary effects, arising from the ways in which the perceived costs and benefits of subsequent educational decisions vary by class origin, lead to further class differentiation of outcomes, even amongst those with similar levels of early achievement. This account has a clear affinity with a description of the form, 'early achievement is necessary but not sufficient for later achievement'. Furthermore, introducing causal heterogeneity, we might hypothesise that, for some classes, early achievement will tend to be necessary but not sufficient, but others, sufficient but not necessary (see Cooper & Glaesser, 2010b, for some relevant evidence).

- Bourdieu's theory of capitals has included, at various times, the claim that educational capital has to be combined with other forms of capital to receive its full economic and social return. Here, educational capital is not sufficient for certain outcomes unless conjoined with other forms such as social capital (Bourdieu, 1974).

Given these affinities, we have argued elsewhere for the use of set theoretic approaches in establishing the regularities relevant to evaluating the claims of such theoretical models as those of Boudon and Bourdieu (Cooper & Glaesser, 2010b). In this paper, we begin the move to the next stage of our work, by using set theoretic QCA to select cases for in-depth study which has the aim of increasing knowledge of the causal processes and mechanisms generating these regularities.

Linear regression / QCA

Before discussing the approach to case selection we have adopted, we should note the crucial difference between the nature of the regularities produced by, on the one hand, simple regression approaches and, on the other, by QCA. Notwithstanding the existence of more complex forms of regression (those including interaction terms, dummy variables, etc.) the basic form of the equations produced by the linear correlational approach have the additive form, for two independent variables X and Z, and for case i:

$$Y_i = \text{Constant} + AX_i + BZ_i + e_i \text{ (an error term).}$$

Ragin has noted, of such simple forms of regression:

> In the net-effects approach, estimates of the effects of independent variables are based on the assumption that each variable, by itself, is capable of producing or influencing the level or probability of the outcome. While it is common to treat 'causal' and 'independent' as synonymous modifiers of the word 'variable', the core meaning of 'independent' is this notion of autonomous capacity. Specifically, each independent variable is assumed to be capable of influencing the level or probability of the outcome *regardless of the values or levels of other variables* (i.e., regardless of the varied contexts defined by these variables). Estimates of net effects thus assume *additivity*, that the net impact of a given independent variable on the outcome is the same across all the values of the other independent variables and their different combinations. (2006: 14-15, emphasis in original)

In contrast, QCA, employing the concepts of necessary and sufficient conditions, is a holistic approach, orientated to analysing causal heterogeneity. QCA drops the assumption of the independence of causes, preserving during analysis the particular combinations of features of a case. QCA, instead of talking about independent and dependent variables, usually refers to conditions and outcomes.

Mahoney and Goertz (2006) offer this invented illustrative example of a Boolean equation:

$$Y = (A*B*c) + (A*C*D*E)$$

Here, capital letters stand for the presence of a condition, lower case letters for its absence, the asterisk symbol denotes logical AND (set intersection), and the plus symbol logical OR (set union). In this example, there are two alternative pathways to the outcome Y: either the joint presence of conditions A and B, coupled with the absence of C, or the joint presence of A, C, D and E. Either pathway is sufficient for the outcome to occur, but neither is necessary, given the existence of the other.

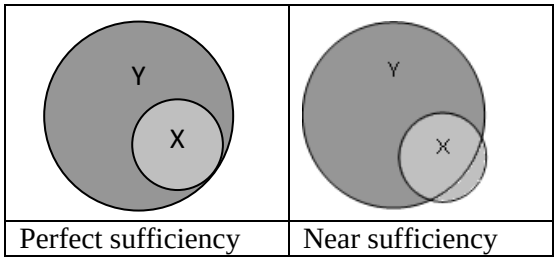
Necessary and sufficient conditions involve relations of logical implication, as shown, for perfect sufficiency, in Table 1 (based on Boudon 1974b; as discussed by Cooper 2005).

Table 1: Sufficient conditions

'if X, then Y', a sufficient relationship		
	Y	Not Y
X	Present	Excluded
Not X	Possible	Possible

This relationship of perfect logical sufficiency is reflected in the Venn diagram in the left hand panel of Figure 1 where X is a sufficient condition for Y, that is, whenever X is present, Y occurs. In set theoretic terms, the cases with the condition form a subset of the cases with the outcome.

Figure 1: Venn diagrams: logical sufficiency



Perfect sufficiency | Near sufficiency

In the social world, relations are less clear-cut than this and we are more likely to find a situation like that depicted in the right-hand panel of Figure 1, where most but not all cases with the condition obtain the outcome. Here X is a condition for Y that may be considered 'nearly always sufficient' or quasi-sufficient. In this framework of fuzzy rather than conventional logic, the proportion of cases with the condition who obtain the outcome provides a measure of the degree of consistency with sufficiency[6]. A consistency of 1 indicates the limiting case of perfect sufficiency. Values of at least 0.7 are usually considered to indicate quasi-sufficiency, though it is open to the analyst to impose a higher, and thus stricter, threshold.

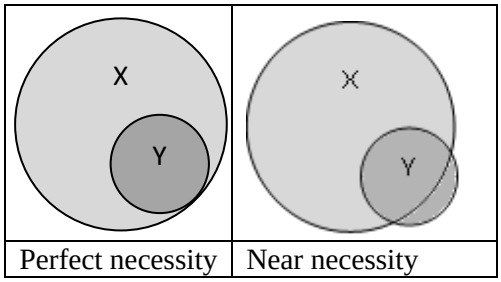
These Venn diagrams can also introduce the concept of explanatory coverage. Analogous to variance explained in regression analysis, coverage indicates how important a condition is with respect to the outcome. In the diagrams (Figure 1), we can see that there must be other conditions which also lead to the outcome, since a substantial proportion of the outcome set Y is not covered by the condition set, X. Numerically, coverage is expressed as the proportion of cases with the outcome who also have the condition[7].

Necessity also involves subsethood. Here, the outcome must be a subset of the condition (see Table 2 and Figure 2), i.e., without the condition, the outcome is not obtained, though not all cases with the condition need obtain the outcome. Consistency with necessity can be calculated in an analogous manner to sufficiency by using the proportion of the cases with the outcome that have the condition.

Table 2: Necessary conditions

'Y, only if X', a necessary relationship		
	Y	Not Y
X	Present	Possible
Not X	Excluded	Possible

Figure 2: Venn diagrams: logical necessity

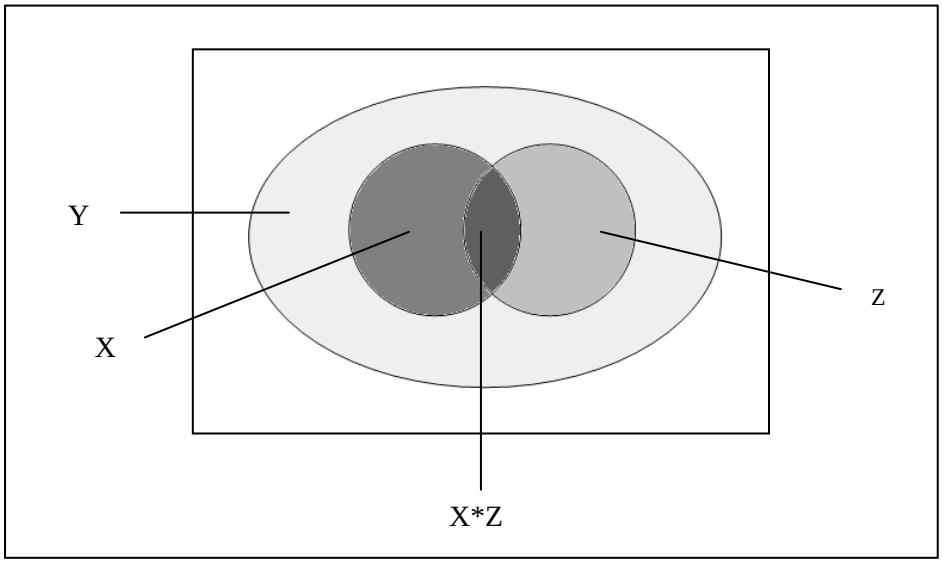


Perfect necessity | Near necessity

We now discuss how QCA deals with several conditions, i.e. the sort of more complex situation in which social science researchers are generally interested (and the one we will ourselves address later in the paper when we explain our process of case selection).

Figure 3 illustrates how two conditions, X and Z, are related, set theoretically, to the outcome Y. They are sufficient conditions because they are subsets of Y, but they are not necessary conditions given that a part of Y is not covered by either of X or Z. The intersection X*Z (the darkest area) comprises cases which have both conditions. The union X + Z, the three darkest areas in Figure 3, contains all those cases who have condition X or condition Z (or both).

Figure 3: Venn diagram: two sufficient conditions



Here, given the relations in Figure 3, the QCA solution would be Y=X+Z. Whenever there are at least two intersecting conditions linked by logical OR, as they are here, explanatory coverage can be partitioned into two components. The unique coverage due to a particular condition is that part of Y which is covered by that

condition but not also by any other sufficient condition(s). In our diagram (Figure 3), the unique coverage of X is that represented by the second darkest area. Raw coverage is the proportion of Y covered by X including its overlap with any other conditions, so here for condition X it is the second darkest and darkest subsets taken together.

If the unique coverage figure for each condition is lower than the raw coverage figure, this indicates that there is some overlap between cases' membership in the condition sets[8]. In fact, a large degree of overlap is not empirically unusual. Taking an example from the sociology of education, we know that parents' education as well as parents' social class can act as conditioning factors for children's educational attainment, and also that high parental education is usually associated with high parental social class. This is a situation that QCA is well suited to addressing, since this latter relationship – or set intersection – is explicitly recognised as an important conjunctural feature of the causal process, rather than being addressed in terms of the larger net effect of one or the other condition.

Later in the paper we draw on and discuss results from our work analysing data from the SOEP (German Socio-Economic Panel). First, here, we will use an initial simple example employing that dataset to illustrate the use of QCA in 'establishing the phenomenon' (Merton, 1987). The SOEP is a representative household panel study, conducted annually since 1984 in West Germany, with East German households added in 1990. The outcome we will employ in this illustration is the type of school young people attended at the age of 17. Of the three hierarchically ordered types of secondary school in Germany, the Gymnasium is the highest in academic status, offering the Abitur which is the qualification allowing university entry, the Realschule is the intermediate type, and the Hauptschule the lowest. In some places, comprehensive schools also exist. At the end of primary school, children receive a 'recommendation' for the type of secondary school they are best suited for, according to their marks and their teacher's assessment. The recommendation is binding to various degrees in most parts of Germany and it takes some effort to enter a higher school type than that recommended by the primary school[9].

For the purposes of this paper, we use all those cases from the SOEP who were born between 1986 and 1990 with no missing values on any of the type of school attended at age 17, parental education, parental social class and recommendation of the primary school (n = 790).

The basis of a QCA analysis is a 'truth table' showing the relationships between types of cases, represented by set intersections of the possible combinations of potentially causal conditions[10], and some outcome set. The factors are coded so that 1 = presence of a condition and 0 = absence of a condition[11]. The two conditions we use in the following example are having at least one parent with the highest school qualification, the Abitur ('ABI_1P'), and having at least one parent in the service class ('SC_1P'). The outcome is whether someone was in Gymnasium at the age of 17. Table 3 is the simple truth table we obtain. Each row shows a configuration of conditions. The first two columns indicate whether the condition is present or absent. The third column labelled 'number' gives the number of cases with this particular combination of conditions. We can see that, for example, there are 55 cases who have at least one parent with Abitur but who don't have a parent in the service class. The last column gives the consistency with sufficiency of each particular configuration, assessing the degree to which the set represented by the particular combination of conditions is a subset of the outcome set. As mentioned above, in the crisp case, this is simply the proportion of cases with the condition (or combination of conditions) who have the outcome. So, for example, out of the 269 cases who have both a parent with Abitur *and* a parent in the service class, 75.8 per cent are in Gymnasium at the age of 17. The rows are shown in descending order of consistency. The 1s and 0s in the outcome column – the penultimate column – have been entered by us after choosing a threshold above which we consider consistency high enough to indicate quasi-sufficiency. The decision here was straightforward, since there is only one row with a consistency above 0.7, and there is a clear drop in consistency from the first to the second row. Therefore only the first row is entered into the Boolean minimisation process, which results in a simple solution with just one pathway leading to the outcome (Table 4). Although the result is easy to derive here, we show, as part of our illustration, the output from the fs/QCA software (Ragin et al. 2006) which can be used to generate both truth tables and solutions.

Table 3: Truth table example

at least one parent with Abitur	at least one parent in the service class	number	in Gymnasium at age 17	consistency
1	1	269	1	0.758
1	0	55	0	0.582
0	1	166	0	0.392
0	0	300	0	0.270

Table 4: QCA output example: quasi-sufficiency, for being in Gymnasium at age 17

	raw coverage	unique coverage	consistency
	---	---	---
ABI_1P*SC_1P	0.534	0.534	0.758
solution coverage: 0.534			
solution consistency: 0.758			

Raw and unique coverage (and coverage of the solution as a whole) are identical here because there is only one quasi-sufficient pathway, that of having at least one parent with Abitur combined with at least one parent in the service class, leading to the outcome. Coverage, at 0.534, is not very low, but it is low enough to indicate that there are many cases following other pathways to the outcome. This pathway is quasi-sufficient but not quasi-necessary.

A process of logical minimisation is used to simplify the solution, where possible, when more than one truth table row, or configuration, passes the chosen threshold for consistency. Purely to illustrate this process, we consider a solution comprising the first two rows of Table 3, i.e. we use a threshold for consistency of 0.55[12]. This gives us the two configurations ABI_1P*SC_1P + ABI_1P*sc_1p, linked by logical OR. Removing logically redundant elements can simplify such a solution. An element is logically redundant if its presence or absence does not make any difference with regard to obtaining the outcome *at the chosen level of consistency*. Here, whether or not a case has a parent in the service class does not make any difference and so we can simplify the solution to just ABI_1P, i.e. having a parent with Abitur. This minimisation process is best undertaken using the fs/QCA software (Ragin et al. 2006) since it becomes complex when more conditions, and therefore more truth table rows, are analysed.

Inspecting the truth table in more detail is worthwhile in itself. Rows, recall, are configurations of conditions or types of cases. This is particularly relevant for our current purposes since we are concerned with selecting cases of a certain type for in-depth study. In this example, we see that the rows containing cases with a more privileged background are found nearer the top of the truth table (which is ordered by consistency), indicating that these groups are more likely to be in the most prestigious school type. Paying attention to the order of the rows can therefore give a first insight into the relations between conditions, their combinations, and the outcome.

Another interesting point to note is that there are fewer cases having just one condition but not the other, i.e. whose parents have Abitur but are not in the service class and vice versa. The last row, the configuration having neither of the conditions, has more cases again. So apart from the relationships between conditions and outcome, this tells us something about how conditions are distributed, and it is not, of course, unusual to find, as we do here, that this distribution is uneven. This arises, of course, from the fact that potential influencing

factors are often related in the social world. This is important to bear in mind because it means that attempting to determine the effect of one particular condition, net of any others that might be present, can be difficult[13]. It also means, in the context of case selection, that analysing cases having one condition but not others which are typically found together with it, can be particularly interesting with a view to exploring possible causal processes. Obviously, in our example here, the lowest number (55) is not very low, but we have used so far only two conditions. However, the number of rows in a truth table rises exponentially with each condition added, so that a truth table with three conditions has $2^3 = 8$ rows, four conditions result in 16 rows and so on, which can lead quickly to very low n or lack of cases altogether in some rows.

Selecting cases to develop causal understanding of regularities

Our final purpose in this paper is to explain how we have used the imperfect regularities produced by prior set theoretic analyses to select cases in order eventually to develop understanding of the causal sources of these regularities. We will not, given this particular purpose and limitations of space, attempt here to undertake a systematic review of the myriad previous approaches to case selection, some of which are oriented to different purposes. However, before we develop a more complex QCA of the conditions predicting a student being in the Gymnasium at age 17, and demonstrate how we have used it to select cases for in-depth study, we will briefly note some relevant recent approaches.

Of course, given that our work begins with set theoretic cross-case analyses of all the cases in some existing dataset meeting some criteria of location, year of birth, etc., we are not in a position to select cases iteratively to develop theoretical knowledge via theoretical sampling (Glaser and Strauss, 1967) or by the sampling techniques used by some of the early advocates of analytic induction. Apart from that, our approach is not an inductive one, since it selects the conditions for our QCA models on the basis of existing knowledge. In selecting recent literature to discuss, we have taken account of these two points, but also of our main purpose in selecting case studies, that of developing our understanding of the processes that generate, and therefore explain, the imperfect regularities documented by our QCAs.

George and Bennett (2005), who argue for within-case process-tracing as a route to causal understanding, explicitly classify types of cases in terms of their potential value for explaining outcomes, as part of the process they term 'typological theorising'. They discuss, amongst others, deviant, most likely, least likely and crucial cases each of which can play a specific role in confirming or disconfirming the existence of some hypothesised causal pathway. Seawright and Gerring (2008) use a similar categorisation of cases, relying upon forms of regression analysis to analyse large datasets and basing subsequent case selection on the results. Depending on the goal of the analysis, different kinds of cases may be selected. They discuss seven ways of selecting cases. In keeping with the underlying regression paradigm, a key criterion is whether a case is an 'onlier' (a low-residual case) or an 'outlier' (high-residual). The former is regarded as *typical* for the phenomenon under study, the latter as *deviant*[14].

We will also employ this key distinction, but in the context, of course, of QCA rather than regression. We will call cases conforming to an established conjunctural regularity typical. In the context of Table 3, typical cases for the first row would be the 76% achieving the outcome. Studying such cases can help confirm or challenge our or others' hypothetical accounts of underlying causal mechanisms and processes. For example, using a large sample, we might find that some particular combination of parental education and employment predicts high educational achievement (an empirical regularity), supporting the view that high achievement is brought about by this particular conjunction of social background factors. However, we cannot simply assume that only a single set of underlying causal processes is involved in generating this regularity. In-depth study of some cases who follow this pattern can tell us more about the resources and processes in the home and in the school which bring it about. We may find that well-educated parents with particular types of employment (those that are not too time-demanding perhaps) read more to their children or help them more with their homework, supporting high academic achievement. We might find alternative supportive processes operating where parents' work is more time-demanding, perhaps involving the buying of support in the form of tutors.

Such a finding might suggest to us that we need to expand our QCA model to take account of finer distinctions between types of work and their demands.

We will refer to cases which do not conform to the empirical regularity being focused on as deviant. In the context of Table 3, deviant cases for the first row would be the 24% not achieving the outcome. They are several possible reasons, of course, why such deviant cases might exist, one of which would be the role of chance and contingency in social life. A key possibility though, as we saw from another angle in discussing typical cases, is that a QCA model is too broad/imprecise to cover all cases, given the complexity of social causation and the heterogeneity of cases. In this regard, studying such deviant cases can also help develop and refine theory (a point also made by Ragin, 1987, 2000, in his writing on QCA). Using the example of set theoretic prediction of educational achievement again, studying cases who achieve more highly than expected, given their combination of social background factors, might reveal that a special talent in association with good support at school can act as a generative causal mechanism that results in some cases having had unexpected outcomes under the original model, thus refining our theory about how educational success is generated. Lareau (2003) provides an example of such in-depth case study. Given the likely small number of such deviant cases, generalisation will, however, be difficult here (Lieberson 1991), unless the findings can be confirmed by a new large scale study, so this will be a heuristic rather than theory confirming exercise[15].

A key general point noted by most recent writers is that case selection aimed at developing causal / explanatory knowledge, as it had to be for Glaser and Strauss as part of grounded theorising, needs to be based on purposive sampling rather than simple random sampling. A researcher wanting to explore causal paths to an outcome might want to undertake in-depth study of a very specific type which is rare in the dataset as a whole case (e.g. an Oxbridge entrant from a working class and ethnic minority background) and random sampling is unlikely to deliver many, or even any, such cases[16].

Our own purpose, as part of a larger programme of work on decision making in the German and English education systems, is the use of interview-based case studies to describe/trace the processes generating the hypothetically causal pathways thrown up by our prior application of QCA to large datasets. We expect this part of our work to contribute to theory development in the sense of confirming, refining or challenging existing explanatory theory. We will now return to this work, showing how and why we have used QCA to select cases for interview.

Using QCA to select cases

As we argued earlier, QCA's focus on conjunctural causation conceived in terms of necessity and sufficiency affords a high degree of potential affinity with causal processes in the social world. A further reason for using it rather than a form of regression analysis as the basis for case selection is that both the configurations in its minimised solutions as well as the configurations that enter the minimisation process from the truth table comprise skeletal types of case. In the simple example of a set theoretic solution used earlier to introduce QCA, focusing on those achieving the outcome of being in the Gymnasium, typical cases would be those with at least one parent with the Abitur and at least one parent in the service class. Deviant cases would be those with these two conditions who do not obtain the outcome. We will now describe the more complex QCA analysis we have actually used to select interviewees, approximately 40 in both of England and Germany. We here add two further conditions to the model that generated our initial illustrative truth table (i.e. to Table 3). These are gender (coded 1 = male, 0 = female) and whether the recommendation at the end of primary school was for Gymnasium, 'GY_REC' (1 = was for Gymnasium, 0 = was for Realschule or Hauptschule)[17]. We now have four conditions, generating a truth table with $2^4 = 16$ rows.

Table 5: Truth table with four conditions

male	at least one parent with Abitur	at least one parent in the service class	recommendation was for Gymnasium	number	in Gymnasium at age 17	consistency
0	1	1	1	96	1	0.917
1	1	1	1	111	1	0.847
1	0	1	1	39	1	0.795
1	1	0	1	19	1	0.789
0	1	0	1	18	1	0.778
0	0	1	1	42	0	0.667
1	0	0	1	47	0	0.660
0	0	0	1	61	0	0.639
0	1	0	0	7	0	0.429
0	1	1	0	34	0	0.382
1	1	1	0	28	0	0.321
0	0	1	0	44	0	0.114
0	0	0	0	93	0	0.075
1	0	0	0	99	0	0.040
1	0	1	0	41	0	0.02
1	1	0	0	11	0	0

As before, it is instructive initially to inspect the truth table. Again, we have ordered its rows by the consistency measure. We find that the upper half contains all those configurations with cases who had received the recommendation for the Gymnasium. We can also see that the cases are now distributed very unevenly across rows. One row has just seven cases, indicating that some combinations of factors appear only rarely[18].

Taking account of the substantial drop in consistency after the fifth row, we employ a consistency threshold of 0.75 for quasi-sufficiency. As with Table 3 above, the 1s and 0s entered in the outcome column reflect this decision. Therefore the first five rows have been entered into the logical minimisation process, generating the solution shown in Table 6.

Table 6: Solution for the analysis with four conditions

	raw coverage	unique coverage	consistency
	----------	----------	----------
ABI_1P*GY_REC+	0.552	0.306	0.865
MALE*SC_1P*GY_REC	0.327	0.081	0.833
solution coverage: 0.634			
solution consistency: 0.855			

We see two (partly overlapping) pathways to the outcome: ABI_1P*GY_REC + MALE*SC_1P*GY_REC, linked by logical OR, i.e. either the combination of having at least one parent with the Abitur and having received a recommendation for Gymnasium at the end of primary school, *or*, for males only, the combination of having had that recommendation coupled with having at least one parent in the service class. The first of these elements, ABI_P*GY_REC, comprises the first two and the fourth and fifth truth table rows, i.e. the combinations 0111, 1111, 1101 and 0101 (where the 1s and 0s refer to the entry against the respective conditions). The second element, MALE*SC_1P*GY_REC, comprises the combinations 1111 and 1011. It can be seen that the overlap is the combination 1111. Consistency with sufficiency is fairly high, both for the solution as a whole (i.e., the two pathways linked by OR) and for each pathway. Coverage for the solution as a whole is fairly high, too, but not so high as to indicate that these two are the only pathways leading to the outcome. It is worth noting that the unique coverage of the second pathway is fairly low which indicates that most of the set representing this pathway overlaps the set representing the other pathway[19]. In other words, many cases who have the conditions indicated by the second pathway will also have those of the first (exactly the sort of situation where case studies of processes are required in order to move to any causal understanding). The recommendation for Gymnasium is part of both pathways in the solution. In spite of the overall solution coverage not being high enough to indicate quasi-necessity, a separate check showed the recommendation itself to be a quasi-necessary condition. (For a discussion of the relation between coverage of the solution as a whole and the quasi-necessity of the constituting conditions, see Glaesser and Cooper 2010.) The set theoretic solution highlights another feature of QCA which makes it a fruitful basis for case selection. It is the finding that there are two quasi-sufficient pathways leading to the outcome, which means that there is more than one type of typical case to explore. Regression analysts, in so far as they focus on a simple onlier/outlier distinction, will tend to miss such causal complexity.[20]

We can now use both the Boolean solution and the truth table itself to identify our typical and deviant cases, specific, of course, to this particular QCA model. Cases with quasi-sufficient combinations of conditions who do actually have the outcome we call *typical with regard to sufficiency*. Cases with the conditions who do not have the outcome, we term *deviant with regard to sufficiency*. Cases who are *deviant with regard to necessity* do *not* have a quasi-necessary condition but obtain the outcome nevertheless.

First, as we have argued, studying typical cases can help confirm (or challenge) our hypothetical accounts of underlying causal mechanisms. In our example, given the finding that the conjunction of factors pertaining to parental education and recommendation for Gymnasium is a quasi-sufficient condition for being in Gymnasium at age 17, we would expect that both support in the home and sponsorship through the system act together to maintain an academic educational career. Studying a case with these characteristics should be able to inform us about the specific processes by which this is brought about. These processes may be those we expect, confirming our hypothetical account of the processes that underlie this conjunctural path to the outcome. On the other hand, of course, we might be surprised in some cases to find that our hypothetical understanding is challenged by what process-tracing throws up. Second, studying cases which are deviant with regard to sufficiency can alert us to countervailing forces not incorporated in the initial cross-case analysis which prevented the outcome being achieved despite the presence of the usually favourable conditions. Finally, studying cases which are deviant with regard to necessity can alert us to conditions we have omitted which can serve as functional alternatives to the quasi-necessary condition identified in the initial cross-case analysis. Some examples of typical and deviant configurations (cases) are listed in Table 7, using the analysis in Table 5/Table 6.

Table 7: Types of configurations

	Configuration	**Outcome**	**Types of cases**
1	ABI_1P*GY_REC	present	typical
2	MALE*SC_1P*GY_REC	present	typical
3	ABI_1P*GY_REC	absent	deviant with regard to sufficiency
4	MALE*SC_1P*GY_REC	absent	deviant with regard to sufficiency
5	gy_rec	present	deviant with regard to necessity

A further consideration in identifying cases to advance explanatory knowledge concerns how common certain types of cases/pathways are empirically. For example, if we look at the truth table as a whole, concentrating on those obtaining the outcome, we can calculate that, amongst the 382 cases obtaining the outcome, 62% have the condition ABI_1P and 38% don't. This makes row 3, MALE*abi_1p*SC_1P*GY_REC, particularly interesting since here we have cases who lack ABI_1P and yet who fall into a quasi-sufficient configuration. It therefore seems worth investigating cases with this configuration of factors because, for example, they may help tease out just how the lack of any parent with Abitur can, for those with the recommendation, be compensated for by a particular combination of factors – being male and having at least one parent in the service class. In addition, we may find that one or both parents are unusual representatives, in some respect, of the service class, given their lack of high academic qualifications, and this particularity – in ways yet to be documented – in itself may have contributed to their children's academic success.

Another fruitful way of using the truth table itself – rather than the Boolean solution – is to identify configurations where the consistency is neither near 0 nor near 1, i.e. to focus on what are termed contradictory cases in the QCA literature. To recall, these are sets of cases which show no clear tendency with regard either to obtaining or to not obtaining the outcome. Assuming we are not faced with some inherently random process, this indicates that there must be additional factors which determine which way the case goes, and studying cases with such configurations, both with and without the outcome, can therefore, by identifying such factors, contribute to theory development. Ragin, in his 1987 book, discusses the way in which contradictory rows indicate the need to look for further causal factors, and he suggests following the lead of case-oriented researchers: 'when case-oriented researchers are confronted with inconsistencies or paradoxes comparable to contradictory rows, they typically examine the troublesome cases in greater detail and attempt to identify omitted causal variables' (p.113).[21]

We should point out that the configurations listed in Table 7 and discussed here are by no means the only ones that can be identified as being interesting for further study, for other purposes than ours, from the QCA analysis. Our specific aim has been to illustrate how QCA may be used to inform the case selection process so that, via subsequent in-depth case studies, it can better contribute to theory development.

One important aspect of the approach we have employed concerns equifinality, as a situation with more than one causal pathway to an outcome is sometimes termed. In our example there are two alternative quasi-sufficient pathways to obtaining the outcome. QCA is well suited to detect such equifinality. In the complex social world, an event or outcome might have such multiple causes[22] and a single cause might have multiple effects as a consequence of the other causes with which it is conjoined. Applied to our purpose, this means that in selecting cases, using QCA ensures that we choose representatives for each hypothetically causal pathway to the outcome focussed on, rather than choosing cases which conform to just one. In Table 7, we recognise this equifinality by including each component of the solution (rows 1 and 2).

A further point concerns the practicalities of using the approach presented here. Our running example used data from the SOEP which is a large anonymised survey not run by ourselves. In fact, we cannot interview the actual respondents from the SOEP since their anonymity has to be preserved. Instead, we are approaching other young people of the same age who show the combinations of characteristics identified through the QCA analysis of the SOEP data as being theoretically interesting cases for more detailed study. In other words, we use QCA to identify types of cases which might be interesting from the point of view of developing our explanatory understanding, and then locate and study cases which conform to those types.

Conclusions and outlook

We have demonstrated how both the truth table and the Boolean solution may be used to inform the case selection process. Using the truth table as well as the solution is important given the minimised nature of the Boolean solution which may, as noted earlier, hide some detail. The truth table, apart from its role in

identifying heterogeneity, is also a useful tool because it maps out all the potential configurations, including those which might be empirically rare, but nevertheless interesting to explore, in order to develop explanatory knowledge. Other researchers have, of course, noted the usefulness of typologies as a means of systematising their selection of cases for particular purposes. Becker (1998) discusses three approaches which, he argues, are forms of truth table analysis. These are Lazarsfeld's Property Space Analysis, Analytic Induction, and Ragin's (1987) Qualitative Comparative Analysis. These differ in important ways, but they have in common that they make it possible to be 'more formal about the requirement to sample for the full range of possibilities' (Becker 1998: 213). In so doing, researchers ensure that no possible configuration of characteristics characterising some cases is overlooked, and so lessen the danger of generalising inappropriately from maybe just one or some possible configurations to all others. This systematic use of typologies can be a useful exercise in itself since it alerts the researcher to the possible existence of cases characterised by unusual combinations of features. Studying such cases may then advance explanatory understanding of complex empirical regularities and their relations with various outcomes.

We illustrated earlier how one of these typological approaches, QCA, given real social data, is used to identify logically quasi-sufficient (and/or quasi-necessary) conditions for some outcome, thereby establishing hypothetical causal pathways to the outcome. The question then arises of *why* they are quasi-sufficient conditions, i.e. we are interested, as a second stage, in the underlying causal mechanisms producing these set theoretic relationships. While QCA (and any other method for large n cross-case analysis) will detect empirical regularities which may reflect underlying causal mechanisms and which, given prior theoretical knowledge, can sometimes be interpreted causally by drawing on such knowledge, at least plausibly, the latter will not always be so. Some findings are always likely to remain theoretically puzzling, especially once we move away from the study of average net effects and focus, configurationally, on a range of types of case. This possibility has motivated our exploration of how QCA, which is built around a recognition of the likelihood of causal heterogeneity, may be used to select cases for detailed study to aid the further development of explanatory theory.

If appropriate, the findings from the chosen case studies might then be used to refine further the initial QCA analysis. This possibility depends on whether there are enough commonalities between the cases studied in depth, but also on whether any new conditions identified as causally important are actually available in the QCA-analysed large dataset (or, less likely, via further data collection). In our example, if we find, via interviews with our selected cases, that grandparents' education seems to be an important factor in differentiating whether or not someone with not very highly qualified parents is in Gymnasium at the age of 17, we will be able, for most cases, to incorporate this into a more detailed QCA of SOEP data since this information can be found in the dataset. It will be another matter if the in-depth case studies point to the causal importance of the extended family, since this information is not available in this dataset.

Using QCA as the method on which to base case selection has several advantages, for those wishing to advance case-based typological approaches, over regression-based approaches to case selection. Given its focus on configurations, the results translate immediately into cases, since cases are conceived of as combinations of characteristics. Using the truth table in addition to the solution is a systematic way of mapping out all the possible combinations of factors. This ensures that no logically possible constellation will be overlooked, even if cases with this configuration of factors turn out to be empirically rare or non-existent. Incidentally, by drawing attention to such configurations, QCA can also inform one's thinking about the topic in hand. Applied to our example, we might explore why there are only relatively few cases where at least one parent has Abitur but where neither parent is in the service class and how this particular constellation influences a child's educational trajectory. These are parents who for some reason have not converted their qualification into an appropriate position in the labour market. This might influence their offspring's schooling in various ways: on the one hand, since the outcome we are looking at concerns educational career rather than social class, we might expect the influence of the parents' high qualification to dominate that of social class, on the other hand, such parents may not place much importance on schooling because their own did not help them get into the service class[23], making them disillusioned with education. Among the teenagers we interviewed in Germany[24], there were two with such a combination of conditions. It turned out that their parents had immigrated to Germany, from Kazakhstan in one case and from Poland in the other, and were

highly supportive of their children's educational careers. In fact, among the reasons for leaving their country of origin had been to provide their children with better lives, and they were ready to forgo higher status jobs they had previously held in order to make this possible.

The approach to case selection we have presented here rests on the idea that we identify *types* of cases for in-depth study. However, our example of the two immigrant families raises a final general problem, that of the extent to which we can safely generalise any findings from such case studies. Cases selected as representatives of their type with regard to the factors under consideration may still be unrepresentative in respect of some condition not included in the initial QCA model, as in the two cases outlined above. It does, of course, seem plausible that enabling one's children to obtain a better education is one reason for immigration, even if it means for the parents that they have to give up status and income, but it may not be possible to establish from a small number of case studies whether all families with the same configuration of parental education and class and immigration status are similarly motivated. In other words, the problem of generalisability re causation persists. If we want to generalise from a typical case, this case has to be similar in relevant characteristics to the cases to which the generalisation is to be extended (Mitchell, 1983), and this may be hard to put into practice. Using QCA in the way we suggest, however, at least ensures that we are aware of the possible relevant combinations of conditions so that we can find cases characterised by each of them in order to explore the range of possible causal processes behind the set theoretic relation between any configuration and the outcome.

Notes

[1] An earlier version of this paper was presented at the British Sociological Association Annual Conference 2010. We would like to thank the two anonymous referees for MIO for their helpful comments on this paper.
This work was supported by an Economic and Social Research Council (ESRC) research fellowship [RES-063-27-0240] awarded to JG.

[2] On the difficulties of this, see Morgan & Winship (2007).

[3] See the usually neglected chapter on quantitative data in Glaser and Strauss (1967).

[4] We have also explored the way in which cluster analysis can be used to simulate set theoretic analysis (Cooper & Glaesser, 2010a), but in this paper we will restrict our discussion to QCA.

[5] It is not only within the quantitative tradition that such an approach has been taken. Analytic induction, for example, aimed to establish invariant relationships between prior factors and some outcome.

[6] This is true in the crisp context, where a case is either completely in or completely out of a set. In the fuzzy context, cases can have partial membership of a set. Since this paper uses crisp sets only, we don't discuss fuzzy sets here but refer the reader to Ragin (2000; 2005; 2008) and also Cooper (2005).

[7] Again, in the crisp context.

[8] The relative size of the condition sets can also be important here.

[9] This is only a very brief overview of the German school system. As this paper has a methodological rather than a substantive focus, we cannot go into any more detail here but refer the reader to Glaesser (2008).

[10] Of course, some rows of this truth table may have no empirical cases.

[11] This reflects our use of crisp rather than fuzzy sets in this paper. We have decided to use crisp rather than fuzzy sets in this particular paper for two reasons. One is that most of the factors we are interested in here lend themselves to dichotomisation with no great loss of information. The other is that configurations and truth tables based on crisp sets translate more straightforwardly into cases which, given our aim of case selection, is particularly important. We are aware, however, of the limitations of crisp set analysis and have explored the use of fuzzy sets in other work of our own (Cooper, 2005; Cooper and Glaesser, 2010a,b).

[12] This wouldn't normally be done because the consistency of the second row is far too low to be considered an indication of quasi-sufficiency. The resulting solution could, of course, be presented as giving the paths that raise the chance of gaining the outcome, ceteris paribus, to over 0.5.

[13] This is the well-known problem, in regression analysis, of multicollinearity.

[14] Their other types of cases, chosen for varying purposes, are diverse, extreme, influential, most similar, and most different cases. Rohlfing (2008) also discusses the use of residuals analysis in case selection. Noting the increasing tendency to combine regression with case study under the banner of mixed methods, he undertakes nested case analysis

(as he calls the combination of large and small n analysis, following Lieberman 2005) specifically to improve the model fit of a regression.

[15] Abell suggests that rather than generalising from the findings concerning one large dataset to other cases, it might be fruitful to study causal processes in individual cases and then determine to what extent they may be generalised (Abell 2009).

[16] One form of theoretically informed sampling of cases for detailed study is to select cases randomly *from within strata* obtained through stratifying on certain variables which are thought to be of theoretical interest (Fearon and Laitin 2008). While this is aimed at obtaining a small number of cases only, the dangers of purely random sampling are mitigated through the stratification process. This deals with the phenomenon which in QCA terms is described as limited diversity (Ragin 1987, 2000), i.e. the fact that not all possible combinations of factors are equally common empirically. Lacey (1970), in his study of a selective grammar school selected individual cases in order to fill all the cells of a table defined by several binary dimensions, effectively employing a set theoretic case selection technique of the sort QCA can support.

[17] Even in areas where comprehensive schools (Gesamtschulen) exist, the recommendation pupils receive is for either Hauptschule, Realschule or Gymnasium.

[18] It should be borne in mind that, apart from any sampling problems, small case numbers may arise because there really are just a few cases with a particular combination of factors, but it could also be the case that actually there aren't any and measurement errors led to their appearing, or, conversely, that without measurement errors there would be more.

[19] There are in fact just 39 cases in the configurational component of the pathway that doesn't overlap (MALE* abi_1p*SC_1P*GY_REC, i.e. 1011), with 31 of these achieving the outcome. In total 382 cases achieve the outcome and these 31 represent 0.081 of these, thus providing the figure in Table 6 for the unique part of the coverage of MALE*SC_1P*GY_REC.

[20] Regression equations can, of course, document multiple paths to an outcome, in so far as, for example, high values on one independent variable can compensate, additively, for low values on another. In a simple case with two independent variables, 2x + 10z might produce the same outcome as 6x + 3z. However, the multiplicity of such alternative paths are not usually related to any notion of types of cases, as they are in QCA.

[21] Obviously, there is also the possibility that the causal factors included in the original model were inappropriate in the first place in which case the whole model would need to be respecified, possibly omitting some of the original factors. However, if factors have been chosen on good theoretical grounds, it seems more likely that the best course of action is to look for additional factors, not leave factors out which have been included for good reasons.

[22] Lindesmith (1981), however, argues that it may be misleading to assume that social causation is complex in this sense. Instead, it is conceivable that the social sciences are not yet advanced enough to clearly differentiate sub-types within an apparently homogeneous outcome. This, in his opinion, is one of the reasons why we might think, incorrectly, that multiple causes exist for the same phenomenon. The same phenomenon may actually be several phenomena, as, illustratively, in types of suicide, each with its own cause, versus simple 'suicide' with, but only apparently, multiple causes.

[23] Of course, we might find some didn't seek entry into this class, for other reasons.

[24] These interviews are, at the time of writing, still being transcribed and await a full analysis.

References

Abell, P. (2009) 'History, case studies, statistics, and causal inference', *European Sociological Review* 25(5) 561-567.

Becker, H.S. (1998) *Tricks of the trade. How to think about your research while you're doing it*, Chicago: University of Chicago Press.

Bhaskar, R. (1978) *A Realist Theory of Science*, Sussex: Harvester Press.

Boudon, R. (1974a) *Education, Opportunity and Social Inequality*. New York: Wiley.

Boudon, R. (1974b) *The Logic of Sociological Explanation*, Harmondsworth: Penguin.

Bourdieu, P. (1974) 'Cultural and social reproduction', in R. Brown (ed.) *Knowledge, Education and Cultural Change*. London: Tavistock.

Cooper (2005) 'Applying Ragin's Crisp and Fuzzy Set QCA to Large Datasets: Social Class and Educational Achievement in the National Child Development Study', *Sociological Research Online* 10(2) < http://www.socresonline.org.uk/10/2/cooper.html >.

Cooper and Glaesser (2008) 'How has Educational Expansion Changed the Necessary and Sufficient

Conditions for Achieving Professional, Managerial and Technical Class Positions in Britain? A Configurational Analysis', *Sociological Research Online* 13(3) < http://www.socresonline.org.uk/13/3/2.html >.

Cooper and Glaesser (2010a) 'Using case-based approaches to analyse large datasets: a comparison of Ragin's fsQCA and fuzzy cluster analysis', *International Journal of Social Research Methodology*: doi: 10.1080/13645579.2010.483079.

Cooper and Glaesser (2010b) 'Contrasting variable-analytic and case-based approaches to the analysis of survey datasets: exploring how achievement varies by ability across configurations of social class and sex', *Methodological Innovations Online* 5(1) 4-23.

Elman, C. (2005) 'Explanatory typologies in qualitative studies of international politics', *International Organization* 59 (Spring) 293-326.

Fearon, J.D. and Laitin, D.D. (2008) 'Integrating qualitative and quantitative methods', in J. M. Box-Steffensmeier, H. E. Brady and D. Collier (eds.) *The Oxford handbook of political methodology*, Oxford: Oxford University Press.

George, A.L. and Bennett, A. (2005) *Case studies and theory development in the social sciences*, Cambridge, Massachusetts: MIT Press.

Glaesser (2008) 'Just how flexible is the German selective secondary school system? A configurational analysis', *International Journal of Research and Method in Education* 31(2) 193-209.

Glaesser and Cooper (2010) 'Employing Ragin's configurational methods to undertake case selection from a large dataset for in-depth study in order to test and develop theory', *BSA Annual Conference*, Glasgow.

Glaser, B.G. and Strauss, A.L. (1967) *The discovery of grounded theory: strategies for qualitative research*, Chicago: Aldine.

Goldthorpe, J.H. (2007a) *On Sociology. Second Edition. Volume One: Critique and Program*, Stanford: Stanford University Press.

Goldthorpe, J.H. (2007b) *On Sociology. Second Edition. Volume Two: Illustration and Retrospect*, Stanford: Stanford University Press.

King, G., Keohane, R.O. and Verba, S. (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton: Princeton University Press.

Lacey, C. (1970) *Hightown Grammar. The school as a social system*, Manchester: Manchester University Press.

Lareau, A. (2003) *Unequal childhoods. Class, race, and family life*, Berkely: University of California Press.

Lieberman, E.S. (2005) 'Nested Analysis as a Mixed-Method Strategy for Comparative Research', *American Political Science Review* 99(3) 435-452.

Lieberson, S. (1991) 'Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases', *Social Forces* 70(2) 307-320.

Lindesmith, A.R. (1981) 'Symbolic interactionism and causality', *Symbolic Interaction* 4 87-96.

Little, D. (1997) '*Agents, Structures, and Social Contingency: New Thinking About the Foundations of the Social Sciences'*, <http://www-personal.umd.umich.edu/~delittle/PAPER%20FOR%20TSINGHUA%20DDM%20LECTURE%20delivered.htm >, accessed 14/06/11.

Mahoney, J. and Goertz, G. (2006) 'A tale of two cultures: contrasting quantitative and qualitative research', *Political Analysis* 14(3) 227-249.

Merton, R.K. (1987) 'Three fragments from a sociologist's notebooks: Establishing the phenomenon, specified ignorance, and strategic research materials', *Annual Review of Sociology*, 13 1-28.

Mitchell, J.C. (1983) 'Case and situation analysis', *Sociological Review*: 187-211.

Morgan, S.L. and Winship, C. (2007) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, Cambridge: Cambridge University Press.

Pawson, R. (1989) *A Measure for Measures: A Manifesto for Empirical Sociology*, London: Routledge.

Pawson, R. (2008) *'*Causality for beginners'**,** *ESRC/NCRM Research Methods Festival*. Available at: < http://eprints.ncrm.ac.uk/245/.> Accessed 14/06/11.

Ragin, C.C. (1987) *The Comparative Method. Moving beyond Qualitative and Quantitative Strategies*, Berkeley, Los Angeles, London: University of California Press.

Ragin, C.C. (2000) *Fuzzy-Set Social Science*, Chicago and London: University of Chicago Press.

Ragin, C.C. (2005) *From Fuzzy Sets to Crisp Truth Tables*. < http://www.compasss.org/files/WPfiles/Raginfztt_April05.pdf>, accessed 14/6/11.

Ragin, C.C. (2006) 'The Limitations of Net-Effects Thinking', in B. Rihoux and H. Grimm (eds) *Innovative Comparative Methods for Policy analysis*, New York: Springer.

Ragin, C.C. (2008) *Redesigning Social Inquiry: Fuzzy Sets and Beyond*, Chicago: University of Chicago Press.

Ragin, C.C., Drass, K.A. and Davey, S. (2006) *Fuzzy-Set/Qualitative Comparative Analysis 2.0.*, Tucson, Arizona: Department of Sociology, University of Arizona. <http://www.u.arizona.edu/%7Ecragin/fsQCA/software.shtml> Accessed 14/6/11.

Rohlfing, I. (2008) 'What you see and what you get. Pitfalls and principles of nested analysis in comparative research', *Comparative Political Studies* 41(11) 1492-1514.

Seawright, J. and Gerring, J. (2008) 'Case selection techniques in case study research. A menu of qualitative and quantitative options', *Political Research Quarterly* 61(2) 294-308.

Biographies:

Judith Glaesser is an ESRC Research Fellow in the School of Education at Durham University, where she will take the post of lecturer in education from October 2011. Her interests include sociology of education, inequality and meritocracy in education, and research methods, particularly QCA. She studied for a PhD at Konstanz University (published as *Soziale und individuelle Einflüsse auf den Erwerb von Bildungsabschlüssen).* Currently, with Barry Cooper, she is exploring the application of case-based methods to large datasets in comparing transitions in the English and German secondary school systems. A new book, Cooper, Glaesser, Gomm and Hammersley's *Challenging the Qualitative-Quantitative Divide: Explorations in Case-focused Causal Analysis* will be published by Continuum in 2012.

Barry Cooper is Emeritus Professor of Education at Durham University where he was, from 1998 to 2005, Director of Research in Education. He was from 2004-2007 co-editor of the *British Educational Research Journal*. His interests are in the sociology of education, especially social class, educational achievement and assessment, set-theoretic research methods and the evaluation of educational aid projects. His most recent book was, with Máiréad Dunne, *Assessing Children's Mathematical Knowledge: Social class, sex and problem-solving.* A new book, Cooper, Glaesser, Gomm and Hammersley's *Challenging the Qualitative-Quantitative Divide: Explorations in Case-focused Causal Analysis* will be published by Continuum in 2012.